

令和 4 年 6 月 30 日現在

機関番号：82508

研究種目：基盤研究(C) (一般)

研究期間：2019～2021

課題番号：19K06315

研究課題名(和文)セマンティックウェブ技術を用いた遺伝素材の交配・解析システムの選定を支援するシステム

研究課題名(英文)Applying semantic web technology to information of trait data in plant varieties.

研究代表者

市原 寿子 (ICHIHARA, HISAKO)

公益財団法人かずさDNA研究所・先端研究開発部・特任研究員

研究者番号：50362398

交付決定額(研究期間全体)：(直接経費) 2,400,000円

研究成果の概要(和文)：交配や解析に使用するシステムの組み合わせの選定は、育種担当者の経験知に基づいて行われてきたが、全てのシステムに関する知識を備えるのは困難であり、また選定するシステムに偏りが生じる可能性がある。本研究では、蓄積されてきた遺伝素材の「システムの親子関係」や「各システムが示す形質の情報」にセマンティックウェブ技術を適用することで「階層的で複数世代に渡る系譜情報の自動的な構築」「形質情報を介したシステム間の共通点・相違点の効率的な抽出」を実現し、育種研究者が交配や解析に適する組み合わせ候補の選定を支援するシステムの構築を目指した。最終的に候補システムの情報抽出は可能になったが、より適切な候補を絞込む基準の導入が必要であると考えられた。

研究成果の学術的意義や社会的意義

本研究では、構築される階層的な系譜情報に基づいて網羅的なシステム情報を抽出できる。これを辿ると複数世代前のシステムや遠縁にあたるシステムの形質情報を一斉に抽出でき、あるシステムの着目形質について系譜の中で表現型として現れない世代があったとしても、由来する祖先システムの全ての候補を提示することが可能である。こうした情報の網羅性は人間が不得意とするところであり、適切な育種親や解析システムの選定上、重要な手掛かりとなる。一部の作物では、品種名や育成者の情報がセマンティックウェブに対応した形式で公開されており、本研究の情報との連結によって、作物流通における生産者・消費者のデータ収集を通じて、経済活動への波及効果が期待できる。

研究成果の概要(英文)：Lined for breeding and analysis has traditionally been selected based on the experimental knowledge of the breeding specialist. However, it is difficult for even specialists to know all candidate lines, and there is a possibility of bias in selecting them. This study applied semantic technology to the "relationship between parent and offspring of lines" and "trait information that each line presents" of the genetic material accumulated so far. Finally, it was possible to extract information on candidate lines.

研究分野：農業情報

キーワード：セマンティックウェブ 作物 育種 形質 系譜 RDF

## 様式 C - 19、F - 19 - 1、Z - 19 (共通)

### 1. 研究開始当初の背景

食料の安定供給や気候変動に対応した系統(または品種) 国際的な食品の流通に対応した食品開発用の系統、および機能性に優れた系統などの育成が喫緊の課題であり、育種の効率化と高速化が求められている。育種の効率化を図る上で、着目形質をもつ系統の創出を目的とした交配親の組み合わせの選定や、着目形質を支配する候補遺伝子の同定を目的としたゲノムワイド関連解析(以下、GWAS)において解析対象とする系統群の選定は非常に重要である。では「複数の着目形質をもつ後代を得るのに適切な両親の組み合わせは何か?」では「着目する形質を支配するゲノム上の多型候補を抽出する上で、その形質に多様性があるに比較するのに最適な系統群は何と何か?」という観点で、系統群の組み合わせを選定する必要がある。

近年、機械学習によりゲノム情報と形質データから遺伝的能力を予測し、優良個体を効率的に選抜するゲノミックセレクション(以下、GS)の有効性が示されているが、はGSに先んじて行われる。従来、交配や解析に使用する系統の組み合わせの選定は育種担当者の経験知に基づいて行われてきたが、全ての系統に関する知識を備えるのは困難であり、また選定する系統に偏りが生じる可能性がある。

こうした背景から、育種研究者が着目する系統や形質の情報に基づいて、交配や解析に適する組み合わせ候補を偏りなく抽出、提示して選定を支援するシステムが求められた。

### 2. 研究の目的

多数の系統および形質を同時に比較できるシステムを構築し、育種研究者が着目する系統や形質の情報に基づいて、交配や解析に適する組み合わせ候補を提示して選定を支援するシステムの構築を目的とした。公開されている遺伝素材の「系統の親子関係」や「各系統が示す形質の情報」に対して、特に、セマンティック技術を適用することによって、「1. 階層的で複数世代に渡る系譜情報の自動的な構築」、「2. 形質情報を介した系統間の共通点・相違点の効率的な抽出」の実現を期待した。セマンティックウェブ技術では、ウェブ上の様々なリソースの内容や関係性をコンピュータが識別できるように定義し、共通する識別子を介して情報の連結・判別・抽出を自動化させることができる。

### 3. 研究の方法

1. 育種素材となる作物の系統・形質の情報を収集し、(1)「系統間の親子関係」と(2)「各系統が保有する形質情報」をセマンティックウェブの標準的な記述形式(図(a) Resource Description Framework、以下、RDF)に変換した。

情報は、公共データベース(以下、DB)である農林水産省品種登録ホームページ(<http://www.hinshu2.maff.go.jp/>、以下、品種登録 HP)や「農業生物資源ジーンバンク(<https://www.gene.affrc.go.jp/>、以下、ジーンバンク)」、各県の農業試験場の紀要、学術論文から取得した。

品種登録 HP の形質情報と形質値は、人間が理解し易い「自然言語」で記述されており、そのままとコンピュータによる処理が困難なため、必要な用語と値を抽出した。

収集、抽出した情報を RDF に変換した。

セマンティックウェブの分野において、作物の名称、関係性や形質値などを説明する語彙や識別子が共通用語として整理された用語集(オントロジー)が公開されているので、変換に際しては、これらを参照し実施した。

2. システムを運用し公開するためのサーバーと DB 設置のためのシステム構築を構築した。ウェブサーバーを設置した。

RDF データを管理するためのプログラム Virtuoso、(<https://virtuoso.openlinksw.com/>)を採用して、デポジトリを設置した。

1で生成した RDF データを、設置したデポジトリに格納した。この時に、デポジトリの運用プログラムである Virtuoso の機能によって情報間で共通する識別子を介してデータが連結され、系統と形質の情報が自動的に高度に階層化される。

Virtuoso は、特殊な検索言語による検索インターフェースも擁しているため、検索を行い結果や検索スピードを検証する。検索スピードが期待したよりも遅い場合は、1- の情報同士の関係あるいは 1- の使用する語彙を再考し、情報の繋がり方を見直した。

### 4. 研究成果

品種登録HPおよびジーンバンクから公開されている作物の150品目の約31,000件のデータを対象として、データセットとデータ検索システムを構築した。この時、登録品種数や形質の観測値が極端に少ない品目については、大量の空データが生成されシステムの検証効率を下げる原因と考えられた。このため、情報量が多い作物を選定して作業を実施した。特に、全国で育成が推奨されている政令指定作物は登録数および観測値に関する情報量が多かったため、これらを中心としたデータ生成を実施した。検索インターフェースについては、当初の計画ではJavaScriptによるグラフィカル・ユーザー・インターフェース(以下、GUI)を目指していたが、検索言語SPARQLの知識を必要とするインターフェースの提供にとどまる結果となった。

本システムでは、元来、独立したデータベースである品種登録HPとジーンバンクのそれぞれの

エントリーを一括で検索、抽出できるようになった。検索例を挙げると、果樹のリンゴ品種名「秋映」から、直接親子関係になく2世代以上離れた母方(種子親)の曾祖父系統「デリシャス」をはじめ、祖父系統「ふじ」、祖母系統「東光」の発芽期、満開期、糖度などの形質値や、観測地、観測年などの観測条件の情報を検出できるようになった。また、病害の一つである「心かびの発生」について、品種登録HPとジーンバンクのそれぞれに由来する情報について、品種単位での頻度、度合い、観測日を抽出できるようになった。最終的に交配親や解析系統候補の情報抽出は可能になったが、情報量が多く、今後、さらに適切な候補を絞り込むための基準の導入が必要であると考えられた。

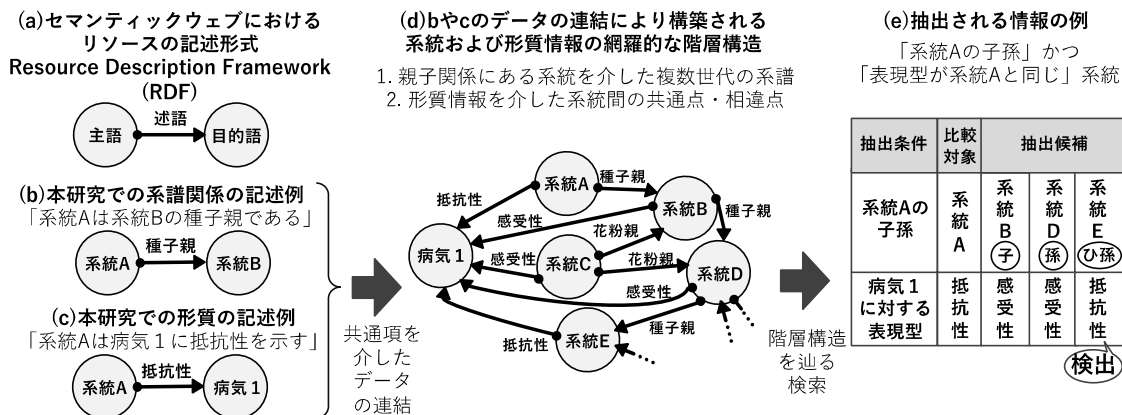


図. 生成されたデータセットの構造と抽出された情報の例

5. 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計2件（うち招待講演 0件 / うち国際学会 0件）

1. 発表者名 市原寿子1、藤井浩2、磯部祥子3、櫛田達矢4,5、田畑哲之3、中谷明弘1（1. 大阪大学大学院医学系研究科、2. 農研機構果樹茶業研究部門、3. かずさDNA研究所、4. JST・NBDC、5. 理化学研究所バイオリソース研究センター）
2. 発表標題 作物品種の形質情報に対するセマンティックウェブ技術の適用 (Applying semantic web technology to information of trait data in plant varieties.)
3. 学会等名 第42回日本分子生物学会年会、口頭発表(公募ワークショップ：フィールドに立ち向かう植物科学、演題番号3PW-07-3、12月5日)、ポスター発表(演題番号4P-0674、12月6日)、福岡
4. 発表年 2019年

1. 発表者名 市原寿子1、藤井浩2、磯部祥子3、櫛田達矢4、田畑哲之3、中谷明弘1（1. 大阪大学大学院医学系研究科、2. 農研機構果樹茶業研究部門、3. かずさDNA研究所、4. JST・NBDC）
2. 発表標題 セマンティックウェブ技術による果樹の系統および形質の情報検索システムの構築
3. 学会等名 令和元年度果樹バイテク研究会、ポスター発表、10月9日 - 10日、和歌山
4. 発表年 2019年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究協力者	藤井 浩  (Fujii Hiroshi)  (00355398)	静岡大学・農学部・客員教授    (13801)	

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------